ELSEVIER

Theoretical Computer Science 255 (2001) 667–677

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

On Parikh slender context-free languages

Juha Honkala[a,b,*,1]

[a] *Department of Mathematics, University of Turku, FIN-20014 Turku, Finland*
[b] *Turku Centre for Computer Science (TUCS), FIN-20520 Turku, Finland*

Received October 1997; revised August 2000; accepted August 2000
Communicated by G. Rozenberg

Abstract

In a recent paper we defined and studied Parikh slender languages and showed that they can be used in simplifying ambiguity proofs of context-free languages. In this paper Parikh slender context-free languages are characterized. The characterization has diverse applications. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Context-free languages; Parikh slender languages; Bounded languages

1. Introduction

Length considerations are often useful in language theory. For example, Flajolet [5] has shown that the inherent ambiguity of many context-free languages can be deduced from the transcendentality of their generating functions. Other deep results based on length considerations are well known, e.g., in the theory of Lindenmayer systems (see [18]).

Andraşiu et al. [1] have defined and studied languages with the property that for each n the number of words in the language of length n is bounded from above by a constant. They have termed such languages *slender*. By now the theory of slender languages has been developed in many directions in Păun and Salomaa [14–16], Dassow et al. [3], Ilie [11], Raz [17] and Nishida, Salomaa [13]. We mention only that slender languages are also of cryptographic interest.

* Correspondence address: University of Turku, Department of Mathematics, 20014 Turku, Finland.
E-mail address: juha.honkala@utu.fi (J. Honkala).

1 Research supported by the Academy of Finland.

PII: S0304-3975(00)00393-5

As a generalization of the approach of Andraşiu et al. [1] the notion of a Parikh slender language was introduced in Honkala [8]. Instead of words of length n we count the number of words with the same Parikh vector. A language L is termed *Parikh slender* if there is a positive integer k such that there does not exist more than k words in L having the same Parikh vector. For basic results concerning Parikh slender languages see Honkala [8], where it is also shown that Parikh slender languages (and power series) can be used in ambiguity proofs of context-free languages. In particular, a new simple proof of the result of Autebert et al. [2] concerning the inherent ambiguity of coprefix languages of infinite words is given.

In this paper we characterize Parikh slender context-free languages. Standard terminology and notation concerning formal languages will be used. Whenever necessary, the reader may consult Ginsburg [6] and Salomaa [19]. We now outline the contents of the paper.

Section 2 contains the basic definitions. We also recall some earlier results. In Section 3 we define Dyck loop languages and establish their connections to bounded context-free languages. Section 4 contains the characterization of Parikh slender context-free languages. As a corollary we obtain a new proof of the result of Ilie [11] and Raz [17] characterizing slender context-free languages.

2. Definitions and previous results

Consider a language L over the alphabet Σ. L is said to be *thin* if for some n_0,

$$\mathrm{card}(\{w \in L \mid |w| = n\}) \leqslant 1 \quad \text{whenever } n \geqslant n_0.$$

L is said to be *slender* if there exists a positive integer k such that

$$\mathrm{card}(\{w \in L \mid |w| = n\}) \leqslant k \quad \text{for all } n \geqslant 0.$$

The definitions of thin and slender languages are due to Andraşiu et al. [1].

If $\Sigma = \{a_1, \ldots, a_m\}$ is a finite alphabet and $w \in \Sigma^*$ is a word, the Parikh vector $\psi(w)$ of w is defined by

$$\psi(w) = (\#_{a_1}(w), \ldots, \#_{a_m}(w)),$$

where $\#_a(w)$ is the number of the occurrences of the letter a in w. Now, a language $L \subseteq \Sigma^*$ is termed *Parikh thin* if for almost all $(i_1, \ldots, i_m) \in \mathbf{N}^m$ there is at most one word $w \in L$ with the Parikh vector $(i_1, \ldots, i_m)$. Furthermore, a language $L \subseteq \Sigma^*$ is termed *Parikh slender* if there is a positive integer k such that for each $(i_1, \ldots, i_m) \in \mathbf{N}^m$ there are at most k words in L with the Parikh vector $(i_1, \ldots, i_m)$. (Here $\mathbf{N}$ is the set of nonnegative integers.)

The following notions are used in the characterization of slender languages. A language $L \subseteq \Sigma^*$ is said to be a *union of single loops* (briefly, USL) if for some k and words $u_i, v_i, w_i \in \Sigma^*$,

$$L = \bigcup_{i=1}^{k} u_i v_i^* w_i.$$

$L \subseteq \Sigma^*$ is said to be a *union of paired loops* (UPL) if for some k and words $u_i, v_i, w_i, x_i, y_i \in \Sigma^*$,

$$L = \bigcup_{i=1}^{k} \{u_i v_i^n w_i x_i^n y_i \mid n \geqslant 0\}.$$

The following result was established in Păun and Salomaa [16] (see also [20]).

Theorem 1. *A regular language L is slender if and only if L is USL.*

The next result was conjectured in Păun and Salomaa [16] and proved by Ilie [11] and Raz [17].

Theorem 2. *A context-free language L is slender if and only if L is UPL.*

Raz [17] also shows that it is decidable whether or not a given context-free language is slender.

We next recall the characterization of Parikh slender regular languages.

A language $L \subseteq \Sigma^*$ is said to be a *multiple loop language* if there exist $k \geqslant 0$ and $u_1, v_1, u_2, v_2, \ldots, u_k, v_k, u_{k+1} \in \Sigma^*$ such that

$$L = u_1 v_1^* u_2 v_2^* u_3 \ldots u_k v_k^* u_{k+1} \tag{1}$$

and

$$\psi(v_1), \ldots, \psi(v_k) \text{ are linearly independent over } \mathbf{Q}. \tag{2}$$

A language $L \subseteq \Sigma^*$ is said to be a *union of multiple loops* (UML) if L is a finite disjoint union of multiple loop languages. Note that if (1) and (2) hold and $w \in L$ there exist unique integers $i_1, \ldots, i_k$ such that

$$w = u_1 v_1^{i_1} u_2 v_2^{i_2} u_3 \ldots u_k v_k^{i_k} u_{k+1}.$$

The following result was shown in [8].

Theorem 3. *A regular language L is Parikh slender if and only if L is UML.*

The characterization of Parikh slender context-free languages is based on the following result due to Honkala [8].

Theorem 4. *Each Parikh slender context-free language is bounded.*

Recall that a language $L \subseteq \Sigma^*$ is said to be *bounded* if there exist words $w_1, w_2, \ldots, w_n \in \Sigma^*$ such that $L \subseteq w_1^* w_2^* \ldots w_n^*$. We need also the following characterization of bounded context-free languages from Ginsburg and Spanier [7]. If u, v are words and L is a language denote

$$(u,v) * L = \bigcup_{n \geqslant 0} u^n L v^n.$$

Theorem 5. *The family of bounded context-free languages is the smallest family of sets containing all finite sets and closed with respect to the following operations*:
1. *finite union*;
2. *finite product*;
3. $(u,v) * L$, *where* u *and* v *are words.*

3. Dyck loop languages

Let X and Y be finite alphabets with $X \cap Y = \emptyset$, and denote $\bar{X} = \{\bar{x} \mid x \in X\}$. The set D of *modified Dyck words* (shortly, D-words) over $X \cup \bar{X} \cup Y$ is the smallest subset $\mathscr{R}$ of $(X \cup \bar{X} \cup Y)^+$ satisfying the following conditions:
1. $Y \subseteq \mathscr{R}$,
2. if $u, v \in \mathscr{R}$ then $uv \in \mathscr{R}$,
3. if $u \in \mathscr{R}$ and $x \in X$ then $xu\bar{x} \in \mathscr{R}$.

A word $u \in D$ is called a *D-prime* if $u \notin D^2$. Clearly, if $u \in D$, there exist D-primes $u_1, u_2, \ldots, u_n$ such that $u = u_1 u_2 \ldots u_n$.

If $u \in D$ and $u = w_1 w_2$, where $w_1, w_2 \in (X \cup \bar{X} \cup Y)^+$, the number of letters of X in w_1 is greater than or equal to the number of letters of $\bar{X}$ in w_1. It follows that if $u, v \in D$ and $x \in X$ then $xu\bar{x}$ is a prefix of $xv\bar{x}$ only if $u = v$. Hence the set of D-primes is a prefix code. Consequently, each word $u \in D$ can be written as a product of D-primes in a unique way.

Suppose Σ is a finite alphabet and $g : (X \cup \bar{X} \cup Y)^* \to \Sigma^*$ is a morphism. If $u \in D$, the language $L(u, g)$ is defined recursively as follows:
1. if $u \in Y$, then $L(u,g) = \{g(u)\}$,
2. if $u = u_1 u_2 \ldots u_n$ where the u_is are D-primes and $n \geqslant 2$, then

$$L(u,g) = L(u_1,g) L(u_2,g) \ldots L(u_n,g),$$

3. if $u = xv\bar{x}$ where $x \in X$ and $v \in D$, then

$$L(u,g) = \bigcup_{n \geqslant 0} g(x)^n L(v,g) g(\bar{x})^n.$$

By definition, a language $L \subseteq \Sigma^*$ is a *Dyck loop language* (shortly, a DL language) if there exist alphabets $X, \bar{X}, Y$, a D-word u over $X \cup \bar{X} \cup Y$ and a morphism $g : (X \cup \bar{X} \cup Y)^* \to \Sigma^*$ such that

$$L = L(u,g).$$

The *loop sequence* $S(u,g)$ of $L = L(u,g)$ is defined recursively as follows:
1. if $u \in Y$, then $S(u,g)$ is the empty sequence;

2. if $u = u_1u_2 \dots u_n$ where the u_is are D-primes and $n \geqslant 2$, $S(u,g)$ is obtained by catenating the sequences $S(u_1,g), S(u_2,g), \dots, S(u_n,g)$, in this order;
3. if $u = xv\bar{x}$ where $x \in X$ and $v \in D$, then $S(u,g)$ is obtained by catenating the sequences $S(v,g)$ and $(g(x)g(\bar{x}))$, in this order.

By definition, the *loop length* of $L = L(u,g)$ is the length of $S(u,g)$.

Example 1. Let $\Sigma = \{a,b,c\}$. Consider the D-word $u = x_2x_1y_1\overline{x_1x_2}x_3y_2\bar{x}_3$. Define the morphism g by $g(x_1) = a^2$, $g(\bar{x}_1) = ba$, $g(x_2) = cac$, $g(\bar{x}_2) = a$, $g(x_3) = b$, $g(\bar{x}_3) = c$, $g(y_1) = c$, $g(y_2) = \varepsilon$. Then u is the product of D-primes $x_2x_1y_1\overline{x_1x_2}$ and $x_3y_2\bar{x}_3$. Therefore,

$$L(u,g) = \{(cac)^m(a^2)^kc(ba)^ka^mb^nc^n \mid k,m,n \geqslant 0\}$$

and

$$S(u,g) = (a^2ba, caca, bc).$$

Consider a DL language $L = L(u,g) \subseteq \Sigma^*$ with loop length m. Then the mapping $W = W(u,g)$ from $\mathbf{N}^m$ into Σ^* is defined recursively as follows:

1. if $u \in Y$, then $W(\emptyset) = g(u)$,
2. if $u = u_1u_2 \dots u_n$ where the u_is are D-primes and $n \geqslant 2$, then

$$W(k_1,\dots,k_m) = W(u_1,g)(k_1,\dots,k_{j_1})W(u_2,g)(k_{j_1+1},\dots,k_{j_2})\dots W(u_n,g)(k_{j_{n-1}+1},\dots,k_m),$$

   where the sequence $(k_1,\dots,k_m)$ is factorized according to the loop lengths of the $L(u_i,g)$s,
3. if $u = xv\bar{x}$ where $x \in X$ and $v \in D$, then

$$W(k_1,\dots,k_m) = g(x)^{k_m}W(v,g)(k_1,\dots,k_{m-1})g(\bar{x})^{k_m}.$$

Intuitively, $W(k_1,\dots,k_m)$ is the word of L obtained by iterating the ith loop of L k_i times.

Consider a sequence $(e_1,\dots,e_m)$ of elements of $\mathbf{N}^k$. The sequence is said to be *linearly independent* if the sequence contains no vector twice and the set $\{e_1,\dots,e_m\} \subseteq \mathbf{Q}^k$ is a linearly independent subset of the vector space $\mathbf{Q}^k$. (Here $\mathbf{Q}$ is the set of rational numbers.) Otherwise the sequence is said to be *linearly dependent*.

Suppose $L = L(u,g)$ is a DL language with the loop sequence

$$S(u,g) = (w_1,w_2,\dots,w_m). \tag{3}$$

Then the *Parikh loop sequence* $P(u,g)$ of $L(u,g)$ is defined by

$$P(u,g) = (\psi(w_1),\psi(w_2),\dots,\psi(w_m)). \tag{4}$$

The language $L = L(u,g)$ is called a *simple DL language* if the sequence $P(u,g)$ is linearly independent.

Suppose $L = L(u, g)$ is a simple DL language with the loop sequence (3). Then the mapping $W : \mathbf{N}^m \to L$ is a bijection. The surjectivity of W is clear from the definitions. To prove injectivity, suppose $W(s_1, \ldots, s_m) = W(t_1, \ldots, t_m)$, where $s_1, \ldots, s_m, t_1, \ldots, t_m \in \mathbf{N}$. Then

$$s_1\psi(w_1) + \cdots + s_m\psi(w_m) = t_1\psi(w_1) + \cdots + t_m\psi(w_m).$$

Now the linear independence of (4) implies that $s_i = t_i$ for $1 \leqslant i \leqslant m$.

Example 1 (*Continued*). Sequence (4) corresponding to the language considered in Example 1 is

$$((3,1,0),(2,0,2),(0,1,1)),$$

which is linearly independent. The triple (k, m, n) corresponds to the word

$$(cac)^m(a^2)^k c(ba)^k a^m b^n c^n.$$

Next, we discuss the relationship between DL languages and bounded languages.

Theorem 6. *Suppose $L \subseteq \Sigma^*$. Then L is a finite union of DL languages if and only if L is bounded and context-free.*

Proof. Suppose first that $L = L(u, g)$ is a DL language where $u \in (X \cup \bar{X} \cup Y)^+$ and $g : (X \cup \bar{X} \cup Y)^* \to \Sigma^*$ is a morphism. If $u \in Y$, $L(u, g)$ is finite and hence bounded context-free. If $u = u_1u_2 \ldots u_n$ where the u_is are D-primes and the languages $L(u_i, g)$ are bounded and context-free, the language

$$L(u, g) = L(u_1, g)L(u_2, g) \ldots L(u_n, g)$$

is also bounded and context-free. Finally, if $u = xv\bar{x}$ where $x \in X$ and $v \in D$, and $L(v, g)$ is bounded context-free, so is $L(u, g)$ because

$$L(u, g) = (g(x), g(\bar{x})) * L(v, g).$$

Consequently, a DL language is bounded and context-free. Because bounded context-free languages are closed under finite union, also a finite union of DL languages is bounded and context-free.

Suppose then that L is bounded and context-free. We show by an induction following Theorem 5 that L is a finite union of DL languages. First, finite languages are obviously finite unions of DL languages. If $L_1, \ldots, L_n$ are finite unions of DL languages, so is $L_1 \cup \cdots \cup L_n$. To conclude the proof it suffices to show that DL languages are closed under finite product and the operation $*$ considered in Theorem 5. Suppose that $L_1 = L(u_1, g_1), \ldots, L_n = L(u_n, g_n)$ are DL languages. Without loss of generality, we assume that the alphabets of the u_is are pairwise disjoint. Then

$$L_1L_2 \ldots L_n = L(u_1u_2 \ldots u_n, g),$$

where g is the common extension of the g_is. Hence $L_1L_2 \ldots L_n$ is a DL language.

Finally, suppose $w_1, w_2 \in \Sigma^*$ are words. Choose a new letter $x \in X$, denote $u = xu_1\bar{x}$ and extend g_1 by $g_1(x) = w_1$, $g_1(\bar{x}) = w_2$. Then

$$L(u, g_1) = (w_1, w_2) * L(u_1, g_1) = (w_1, w_2) * L_1.$$

Therefore $(w_1, w_2) * L_1$ is a DL language. $\square$

Theorem 7. *Suppose $\Delta = \{a_1, \ldots, a_m\}$ is an alphabet and $L \subseteq a_1^* a_2^* \ldots a_m^*$. Then L is context-free if and only if L is a finite union of simple DL languages.*

Proof. By Theorem 6 a finite union of DL languages is context-free. Conversely, suppose that $L \subseteq a_1^* a_2^* \ldots a_m^*$ is context-free. By Theorem 6, L is a finite union of DL languages. Hence, it suffices to prove that a DL language $L \subseteq a_1^* a_2^* \ldots a_m^*$ is a finite union of simple DL languages.

Suppose $L = L(u, g)$ where $u \in (X \cup \bar{X} \cup Y)^+$ is a D-word and $g : (X \cup \bar{X} \cup Y)^* \to \Sigma^*$ is a morphism. We proceed inductively on the loop length k of $L(u, g)$. First, if $k = 0$, the claim trially holds. Suppose that the claim holds if $k \leqslant t$ and assume that the loop length of $L(u, g)$ equals $t + 1$. Let

$$P(u, g) = (\alpha_1, \ldots, \alpha_{t+1})$$

be the Parikh loop sequence of $L(u, g)$. If $P(u, g)$ is linearly independent, $L(u, g)$ is a simple DL language and the claim is true. Assume then that $P(u, g)$ is linearly dependent. Then there exist $i \in \mathbf{N}$, $1 \leqslant i \leqslant t + 1$ and $p_1, \ldots, p_{t+1}, q_1, \ldots, q_{t+1} \in \mathbf{N}$ such that

$$q_1 \alpha_{p_1} + \cdots + q_i \alpha_{p_i} = q_{i+1} \alpha_{p_{i+1}} + \cdots + q_{t+1} \alpha_{p_{t+1}}, \tag{5}$$

where $\{p_1, \ldots, p_{t+1}\} = \{1, \ldots, t+1\}$ and $p_1 < p_2 < \cdots < p_i$, $p_{i+1} < \cdots < p_{t+1}$ and not all q_js equal zero. Now,

$$L = \{W(i_1, \ldots, i_{t+1}) \mid (i_1, \ldots, i_{t+1}) \in \mathbf{N}^{t+1}\}.$$

Denote

$$L_1 = \bigcup_{1 \leqslant j \leqslant i} \{W(i_1, \ldots, i_{t+1}) \mid (i_1, \ldots, i_{t+1}) \in \mathbf{N}^{t+1} \text{ and } i_{p_j} < q_j\}.$$

We show that $L = L_1$. Trivially $L_1 \subseteq L$. Suppose that $w \in L$. Then there exists $(i_1, \ldots, i_{t+1}) \in \mathbf{N}^{t+1}$ such that

$$w = W(i_1, \ldots, i_{t+1}).$$

By (5) there exist $(s_1, \ldots, s_{t+1}) \in \mathbf{N}^{t+1}$ such that

$$s_1 \alpha_1 + s_2 \alpha_2 + \cdots + s_{t+1} \alpha_{t+1} = i_1 \alpha_1 + i_2 \alpha_2 + \cdots + i_{t+1} \alpha_{t+1}$$

and for some j, $1 \leqslant j \leqslant i$,

$$s_{p_j} < q_j.$$

Hence

$$\psi(W(s_1,\ldots,s_{t+1})) = \psi(W(i_1,\ldots,i_{t+1})) = \psi(w).$$

Because

$$w, W(s_1,\ldots,s_{t+1}) \in a_1^* a_2^* \ldots a_m^*,$$

this implies that $w = W(s_1,\ldots,s_{t+1})$. Hence $w \in L_1$. This concludes the proof of the equality $L = L_1$.

Next, fix an integer j with $1 \leqslant j \leqslant i$. Then

$$\{W(i_1,\ldots,i_{t+1}) \,|\, (i_1,\ldots,i_{t+1}) \in \mathbf{N}^{t+1} \text{ and } i_{p_j} < q_j\}$$
$$= \bigcup_{k=0}^{q_j-1} \{W(i_1,\ldots,i_{t+1}) \,|\, (i_1,\ldots,i_{t+1}) \in \mathbf{N}^{t+1} \text{ and } i_{p_j} = k\},$$

where each term in the union is a DL language with loop length t. Therefore it follows inductively that $L = L_1$ is a finite union of simple DL languages. □

4. Parikh slender context-free languages

Before the characterization of Parikh slender context-free languages we need one lemma.

Lemma 8. *Suppose $L_1 = L(u,g) \subseteq \Delta^*$ is a simple DL language with the loop sequence $S(u,g) = (w_1,\ldots,w_n)$. Assume that $h : \Delta^* \to \Sigma^*$ is a morphism mapping L_1 onto the language $L \subseteq \Sigma^*$ which is injective on L_1. Assume that $\phi : \Sigma^* \to \mathbf{N}^s$ where $s \geqslant 1$, is a monoid morphism. If the sequence $(\phi h(w_1),\ldots,\phi h(w_n))$ is linearly independent, ϕ is injective on L. If the sequence $(\phi h(w_1),\ldots,\phi h(w_n))$ is linearly dependent then for every $t \geqslant 1$ there exist distinct words $y_1,\ldots,y_t \in L$ such that*

$$\phi(y_1) = \cdots = \phi(y_t).$$

Proof. Assume first that the sequence $(\phi h(w_1),\ldots,\phi h(w_n))$ is linearly independent. Suppose $\phi(y_1) = \phi(y_2)$ where $y_1, y_2 \in L$. For $i = 1, 2$, let $u_i \in L_1$ be a word such that $h(u_i) = y_i$. Furthermore, let

$$u_1 = W(i_1,\ldots,i_n)$$

and

$$u_2 = W(j_1,\ldots,j_n),$$

where $(i_1,\ldots,i_n), (j_1,\ldots,j_n) \in \mathbf{N}^n$. Then there is a word $u \in \Delta^*$ such that

$$\phi(y_1) = \phi h(u_1) = \phi h(u) + i_1 \phi h(w_1) + \cdots + i_n \phi h(w_n) \tag{6}$$

and

$$\phi(y_2) = \phi h(u_2) = \phi h(u) + j_1\phi h(w_1) + \cdots + j_n\phi h(w_n). \tag{7}$$

Because $\phi(y_1) = \phi(y_2)$, and by assumption $(\phi h(w_1), \dots, \phi h(w_n))$ is linearly independent, it follows by (6) and (7) that $(i_1, \dots, i_n) = (j_1, \dots, j_n)$. Hence $u_1 = u_2$ and $y_1 = y_2$, which shows that ϕ is injective on L.

Suppose then that $(\phi h(w_1), \dots, \phi h(w_n))$ is linearly dependent. Then there exist distinct n-tuples $(i_1, \dots, i_n)$, $(j_1, \dots, j_n) \in \mathbf{N}^n$ such that

$$i_1\phi h(w_1) + \cdots + i_n\phi h(w_n) = j_1\phi h(w_1) + \cdots + j_n\phi h(w_n).$$

For nonnegative integers $t \geqslant 1$, $0 \leqslant q \leqslant t$, denote

$$\alpha(t, q) = q(i_1, \dots, i_n) + (t - q)(j_1, \dots, j_n).$$

If $0 \leqslant q_1 < q_2 \leqslant t$, clearly $\alpha(t, q_1) \neq \alpha(t, q_2)$. Furthermore, for a fixed $t \geqslant 1$ and any q, $0 \leqslant q \leqslant t$,

$$\begin{aligned}\phi h(W(\alpha(t,q))) &= \phi h(u) + q(i_1\phi h(w_1) + \cdots + i_n\phi h(w_n)) \\ &\quad + (t-q)(j_1\phi h(w_1) + \cdots + j_n\phi h(w_n)) \\ &= \phi h(u) + t(i_1\phi h(w_1) + \cdots + i_n\phi h(w_n)),\end{aligned}$$

where $u \in \Delta^*$ is a word. Hence, if we denote $y_q = h(W(\alpha(t, q)))$ for $0 \leqslant q \leqslant t$, then $y_q \in L$ and

$$\phi(y_1) = \cdots = \phi(y_t). \qquad \square$$

Theorem 9. *Suppose $L \subseteq \Sigma^*$ is context-free. Then L is Parikh slender if and only if L is a finite union of simple DL languages.*

Proof. Suppose $L \subseteq \Sigma^*$ is context-free and Parikh slender. By Theorem 4 there exist words $w_1, w_2, \dots, w_m \in \Sigma^*$ such that

$$L \subseteq w_1^* w_2^* \dots w_m^*.$$

Denote $\Delta = \{a_1, \dots, a_m\}$ and define the morphism $h : \Delta^* \to \Sigma^*$ by $h(a_i) = w_i$, $1 \leqslant i \leqslant m$. By the Cross-Section Theorem due to Eilenberg [4] there exists a rational language $R \subseteq a_1^* a_2^* \dots a_m^*$ such that h maps R bijectively onto $w_1^* w_2^* \dots w_m^*$. Hence h maps $K = h^{-1}(L) \cap R \subseteq a_1^* a_2^* \dots a_m^*$ bijectively onto L. By the closure properties of context-free languages K is context-free. By Theorem 7 there exist simple DL languages $K_1, \dots, K_p$ such that

$$K = K_1 \cup \cdots \cup K_p.$$

Because

$$h(K_1) \cup \cdots \cup h(K_p) = h(K) = L,$$

each $h(K_j)$ is Parikh slender. We show that each $h(K_j)$ is a simple DL language. Fix j, $1 \leqslant j \leqslant p$, and suppose that $K_j = L(u,g)$ has the loop sequence $S(u,g) = (w_1, \ldots, w_n)$. Then $h(K_j) = L(u, hg)$ is a DL language with the loop sequence $(h(w_1), \ldots, h(w_n))$. By Lemma 8, $(\psi h(w_1), \ldots, \psi h(w_n))$ is linearly independent. Therefore $h(K_j)$ is a simple DL language. This concludes the proof that a Parikh slender context-free language is a finite union of simple DL languages.

Conversely, suppose that $L_0 = L(u,g)$ is a simple DL language with the loop sequence $(w_1, \ldots, w_n)$. We now use Lemma 8 in the simple case where h is the identity mapping and conclude that ψ is injective on L_0. Hence L_0 is Parikh thin. Because a finite union of Parikh thin languages is Parikh slender, it follows that a finite union of simple DL languages is Parikh slender. $\square$

The proof of Theorem 9 implies the following result.

Theorem 10. *Suppose L is a Parikh slender context-free language. Then L is a finite union of Parikh thin context-free languages.*

As a corollary of Theorem 9 we also obtain a new proof of Theorem 2. Indeed, suppose $L \subseteq \Sigma^*$ is a slender context-free language. Then L is Parikh slender and hence a finite union of slender simple DL languages. Consider a slender simple DL language $L_1 = L(u,g)$ with the loop sequence $(w_1, \ldots, w_n)$. Lemma 8 implies that the sequence

$$(|w_1|, \ldots, |w_n|)$$

is linearly independent. Hence $n = 0$ or $n = 1$. Therefore L_1 is a paired loop. Hence L is a finite union of paired loops.

Finally, the decidability of Parikh slenderness for context-free languages can be shown by the ideas used to prove Theorem 9. However, a simpler proof is given in Honkala [9].

References

[1] M. Andraşiu, J. Dassow, G. Păun, A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, Theoret. Comput. Sci. 116 (1993) 339–357.

[2] J.-M. Autebert, P. Flajolet, J. Gabarro, Prefixes of infinite words and ambiguous context-free languages, Inform. Process. Letters 25 (1987) 211–216.

[3] J. Dassow, G. Păun, A. Salomaa, On thinness and slenderness of L languages, EATCS Bull. 49 (1993) 152–158.

[4] S. Eilenberg, Automata, Languages and Machines, Vol. A, Academic Press, New York, 1974.

[5] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. 49 (1987) 283–309.

[6] S. Ginsburg, The Mathematical Theory of Context-Free Languages, McGraw-Hill, New York, 1966.

[7] S. Ginsburg, E.H. Spanier, Bounded ALGOL-like languages, Trans. Amer. Math. Soc. 113 (1964) 333–368.

[8] J. Honkala, On Parikh slender languages and power series, J. Comput. System Sci. 52 (1996) 185–190.

[9] J. Honkala, A decision method for Parikh slenderness of context-free languages, Discrete Appl. Math. 73 (1997) 1–4.

[10] J. Honkala, On N-algebraic Parikh slender power series, J. Univ. Comput. Sci. 3 (1997) 1114–1120.
[11] L. Ilie, On a conjecture about slender context-free languages, Theoret. Comput. Sci. 132 (1994) 427–434.
[12] W. Kuich, A. Salomaa, Semirings, Automata, Languages, Springer, Berlin, 1986.
[13] T. Nishida, A. Salomaa, Slender 0L languages, Theoret. Comput. Sci. 158 (1996) 161–176.
[14] G. Păun, A. Salomaa, Decision problems concerning the thinness of D0L languages, EATCS Bull. 46 (1992) 171–181.
[15] G. Păun, A. Salomaa, Closure properties of slender languages, Theoret. Comput. Sci. 120 (1993) 293–301.
[16] G. Păun, A. Salomaa, Thin and slender languages, Discrete Appl. Math. 61 (1995) 257–270.
[17] D. Raz, Length considerations in context-free languages, Theoret. Comput. Sci. 183 (1997) 21–32.
[18] G. Rozenberg, A. Salomaa, The Mathematical Theory of L Systems, Academic Press, New York, 1980.
[19] A. Salomaa, Formal Languages, Academic Press, New York, 1973.
[20] J. Shallit, Numeration systems, linear recurrences, and regular sets, Inform. and Comput. 113 (1994) 331–347.